NON-HUMAN IDENTITY SERIES · PAPER #5

Auditing Agents Under NIS2, DORA, and the EU AI Act

Compliance by construction, when architecture becomes the audit evidence

Mohamad Amin Hasbini

Independent researcher · Paris, France

Published May 19, 2026

ABSTRACT

Regulators no longer ask whether an agent action happened. They ask the deploying organization to reconstruct who or what authorized it, scoped to what, valid until when, and how the boundary was held across the delegation chain. Application logs answer the first question. Only architecture-produced evidence answers the second. NIS2 Article 21, DORA Articles 6/8/17-19, and EU AI Act Annex IV converge on the same artifact: a cryptographically chained receipt of authority that survives partial-chain compromise.

KEYWORDS — AI agent security, non-human identity, post-quantum cryptography, agent authorization, capability tokens, NIS2, DORA, EU AI Act

CITE AS

Hasbini, M. A. (2026). *Auditing Agents Under NIS2, DORA, and the EU AI Act*: Compliance by construction, when architecture becomes the audit evidence. Non-Human Identity Series, Paper #5.
Available at https://mahasbini.org/papers/05-auditing-agents/
PDF https://mahasbini.org/publications/papers/05-auditing-agents.pdf

Auditing Agents Under NIS2, DORA, and the EU AI Act

Compliance by construction, when architecture becomes the audit evidence

Non-Human Identity Series, Paper 5 of 5

TL;DR

- **The audit problem.** Regulators no longer ask whether an agent action happened. They ask the deploying organization to show who or what authorized it, when, scoped to what, and how the boundary was held across the delegation chain. Application logs answer the first question. Only architecture-produced evidence answers the second. The reconstruction obligation is the audit-grade gap that retrofit instrumentation cannot close.
- **The convergence.** NIS2 Article 21, DORA Articles 6 / 8 / 17-19, and EU AI Act Annex IV name different obligations under different enforcement regimes. They demand the same artifact: a cryptographically chained receipt of authority that survives partial-chain compromise. Three regulations, one architectural answer.
- **The architecture maps onto the obligations.** Capability tokens (Paper #2) become the per-action authorization primitive auditors can interpret. Zero-knowledge verification (Paper #3) satisfies privacy and auditability simultaneously. Attenuation discipline (Paper #4) produces the receipt chain that makes reconstruction a queryable property rather than a coordination exercise. Each architectural element from this series maps onto a specific regulatory obligation.
- **The compliance window is staged but already open.** DORA has applied to financial entities since 17 January 2025. NIS2 transposition deadlines passed on 17 October 2024; national enforcement regimes are now active across Member States. The EU AI Act is scheduled to become applicable on 2 August 2026 with exceptions; stand-alone high-risk AI systems are expected to apply from 2 December 2027, and high-risk AI systems embedded in regulated products from 2 August 2028 (per the 7 May 2026 AI Omnibus political agreement). Enterprises that have not begun architectural alignment by Q3 2026 face exposure under DORA and NIS2 today and under the AI Act high-risk regime from late 2027 onward. The conversation needs to land Monday morning.

An April 2026 examination

On April 14, 2026, a Tier-1 European bank reports a major ICT-related incident to its competent authority under DORA Article 19. The bank's AI-assisted mortgage-underwriting agent had recommended approval on a loan that subsequently triggered a fraud alert from a downstream verification service. Under DORA Articles 17-19 and the incident-reporting technical standards, the authority's first operational question is not only whether the action happened. It is whether the bank can reconstruct the authority chain behind the action: which user authorization initiated the agent flow, what scope held at each delegation hop between the user and the recommendation, and

how the chain integrity survived the transition to the third-party verification service. Under the DORA incident-reporting sequence, the bank may have only hours for the initial notification (within four hours of classification and twenty-four hours of detection), seventy-two hours for the intermediate report, and one month for the final report including root-cause analysis.

The bank's logs confirm the action happened. The application trace shows which microservice processed the request, which downstream services it called, and which response was returned to the customer. None of these answer the competent authority's question. The bank's compliance team produces a narrative reconstruction: a policy memo, an internal audit-trail summary, a screenshot of the agent's decision-explanation panel. The competent authority reads them and asks a sharper question: can the bank produce a cryptographic receipt chain that shows, by construction rather than by operator attestation, that the agent's authority remained bounded by its originating mandate from delegation hop zero to the action that affected the customer?

The bank cannot. The architecture was not built to produce that evidence. Retrofit instrumentation captured what happened, not who authorized what under which scope at which depth. Under DORA Articles 17 through 19, the bank is now in the position of demonstrating compliance through narrative rather than artifact. The competent authority records the finding. Penalties under DORA's national-competent-authority enforcement regimes vary by Member State but include significant administrative measures and, in some jurisdictions, fines tied to a percentage of annual turnover. The bank's remediation path will run through Q3 2026 and intersect with its EU AI Act preparation for the 2 December 2027 stand-alone high-risk-system enforcement deadline. The compliance program is now in remediation under one regime and pre-remediation under another, sixteen months into the first and eighteen months out from the second.

This scenario is not a hypothetical reach. It is the structural consequence of building agent systems where authorization is enforced operationally but never produced as architectural evidence. The regulator's question is not exotic. Under DORA today, NIS2 national regimes today, and EU AI Act high-risk enforcement opening in late 2027, it is the kind of question competent authorities, auditors, and supervisory teams will increasingly ask as agent-mediated systems enter regulated workflows.

> ***Regulators no longer ask whether the action happened. They ask whether the architecture can show, by construction, who authorized it and how the authority was held across the delegation chain. Application logs answer the first question. Only architecture answers the second. Three regulations converge on the same artifact: a cryptographically chained receipt of authority that survives partial-chain compromise.***

Three regulations name different obligations under different enforcement regimes. They demand the same artifact: a regulator-readable receipt chain produced by the architecture, not curated by the operator.

Part I: The audit problem

Why retrofit instrumentation fails

Application logs record that something happened. Audit-grade evidence records who or what authorized the action, scoped to what, valid until when, and how the boundary was held at each delegation hop. The two artifacts look similar in a screenshot. They differ in a property that matters under adversarial audit: operator dependency.

Retrofit logs are operator-attested. The deploying organization produces the log, curates the narrative, and presents the result to the regulator. If the operator's instrumentation is incomplete, the gap is invisible to anyone outside the operator's organization. If a compromise occurred at the application layer, the log may itself reflect the compromised state.

Cryptographic receipts are architecture-attested. Each authorization decision is bound to its originating authority by a cryptographic chain that the regulator can verify without operator cooperation. Partial-chain compromise leaves a defensible artifact: the regulator sees exactly where the chain integrity holds and where it does not. The architecture produces the evidence; the operator does not curate it.

Under NIS2, DORA, and the EU AI Act, operator-attested logs may remain necessary, but they are increasingly insufficient where the regulated question is reconstruction of authority, control, dependency, or incident root

cause. Current ENISA transposition guidance on NIS2, joint technical standards from the European Supervisory Authorities on DORA operational resilience, and emerging EU AI Office implementing acts on Annex IV documentation point in the same direction: evidence artifacts that survive operator-independent inspection.

What a receipt chain contains, concretely. Each receipt in the chain binds a minimum set of fields: originating authority identifier, agent identity, delegated subject, audience, action class, scope, caveats, expiry, parent receipt hash, proof-of-possession binding, policy decision identifier, revocation epoch, issuer signature, and evidence pointer. The architectural property the chain produces is reconstructability: any downstream action can be traced back through these fields to its originating authority, with cryptographic integrity at each hop. The fields and binding mechanics are specified in the AAC Construction Specification (companion to Paper #3).

The reconstruction problem

Three questions regulators ask, in order, when examining an agent-mediated action:

1. *Show me the originating authorization for this specific agent action.*
2. *Show me how the authorization narrowed, or did not narrow, at each delegation hop between the originator and the action.*
3. *Show me what happened to this authority when it was revoked, including the propagation latency across the delegation tree.*

Each question is unanswerable from application logs alone. Each is answerable from a receipt chain produced by the attenuation discipline described in Paper #4. The reconstruction problem is not solved by a better dashboard or by an additional log stream. It is solved by an architectural property: the receipt at each delegation hop is itself the verifiable evidence, queryable by action ID without operator narrative.

The practical implication for compliance programs is that the audit-readiness target shifts from log completeness to chain reconstructability. Log completeness is a coverage metric. Chain reconstructability is an architectural property. The two cannot substitute for each other.

The composition problem

Single-hop audit is not chain-of-delegation audit. Most enterprise audit programs (SOC 2, ISO 27001, ISO 42001, even SOX where applicable to in-scope financial reporting controls) verify per-action evidence. Paper #4 showed why this is structurally insufficient for agent ecosystems where chains run on hundred-millisecond timescales and span asynchronous orchestration boundaries. Paper #5 maps the consequence into the regulator's framework: composition failure becomes chain-level escalation, and chain-level escalation is what regulators will examine first under post-incident reviews.

The composition problem cannot be solved by adding controls at each hop. It requires that the controls compose: that the property "authorization remained bounded across the chain" is verifiable end-to-end without re-evaluating each hop independently. Attenuation discipline produces that compositional property by construction. Operator-attested logs do not, regardless of how many hops they instrument.

The cross-domain problem

Cross-organizational delegation, including business-to-business agent calls, cross-cloud agent flows, and supply-chain agent chains, introduces a trust-boundary problem the single-organization audit framework was never

designed to handle. One side's audit-grade evidence may not be enforceable by the other side's regulator. One side's caveat semantics may not be understood by the other side's policy engine. One side's revocation channel may not reach into the other side's enforcement perimeter within any defensible timeline.

The solution architecture is well-known from financial-services reconciliation, where cross-institutional transactions have required end-to-end auditability for decades. Each side preserves its own receipt segment. The regulator-side reconstructs the full chain by cross-organizational receipt assembly, using cryptographic chain integrity to verify that no segment was altered after the fact. The agent generation extends this pattern. The architectural answer is the same: receipts produced at each side, assembled by the regulator, verified by chain integrity.

What changes in the agent generation is the scale and the timescale. Cross-institutional B2B transactions historically settled in hours or days, with reconciliation cycles measured in weeks. Cross-organizational agent chains settle in seconds, with reconciliation cycles that may never be initiated unless an incident triggers them. The architecture has to produce the evidence continuously, not only when reconciliation is requested.

What the receipt chain does not prove

A receipt chain does not prove model correctness, output fairness, training-data quality, lawful basis for processing under GDPR, absence of bias, or absence of compromise before credential issuance. It proves a narrower but critical property: the authority path behind an agent-mediated action can be reconstructed and integrity-checked after the fact. Other compliance obligations (model evaluation under EU AI Act Article 15, data governance under Article 10, lawful basis under GDPR Article 6) require their own evidence artifacts. The receipt chain is necessary for authorization-provenance compliance; it is not sufficient for the broader compliance posture. Conflating the two understates the remaining work and weakens the architectural case.

Part II: Regulatory mapping

The mapping below is the executive summary; the prose that follows develops each row.

Regime	Key obligations	Evidence artifact
NIS2 Article 21	Access control, supply-chain security, cryptography, asset management, incident handling	Non-human-actor authorization graph + scoped capability receipts
NIS2 Article 23	Incident characterization within 24h / 72h / one-month windows	Queryable chain reconstruction by action ID
DORA Articles 6 + 8	ICT risk management framework + asset and dependency mapping	Agent dependency and delegation graph
DORA Articles 17-19	Incident classification, reporting, root-cause analysis	Receipt chain + out-of-chain revocation evidence
DORA Articles 24-27	Digital operational resilience testing, including TLPT	Receipt chain queryable under simulated escalation
DORA Articles 28-30	ICT third-party risk and subcontracting	Cross-organizational receipt segments, assembled by regulator

Regime	Key obligations	Evidence artifact
EU AI Act Articles 16 / 17 / 72 / 73	Provider obligations, quality management, post-market monitoring, serious incident reporting	Technical documentation + audit receipts + monitoring hooks
EU AI Act Annex IV	Architecture, validation, cybersecurity, monitoring, risk management documentation	Architecture evidence pack including receipt-chain reference

This part is the substantive bulk of the paper. The mapping is concrete: each regulation, the articles that matter, the evidence artifact required, and how the architecture from Papers #1 through #4 satisfies the obligation.

NIS2: critical infrastructure and essential services

NIS2 (Directive EU 2022/2555) is a directive rather than a regulation: it requires Member States to transpose its requirements into national law. Approximately 160,000 entities across the European Union fall within its essential and important entity classifications, drawn from sectors including energy, transport, water, banking, financial market infrastructures, health, drinking water, digital infrastructure, ICT service management, public administration, and space. NIS2 entered into force in January 2023; Member States had until 17 October 2024 to transpose the directive, and NIS1 was repealed from 18 October 2024 onward. Regulated entities now face national implementation regimes derived from the directive, with enforcement intensity varying by Member State through 2025 and 2026.

Three articles bear directly on AI agent deployment in regulated entities.

Article 21 (Risk management measures) requires entities to implement appropriate technical, operational, and organizational measures across an enumerated list that includes access control and human resources security. Both extend to non-human actors when an agent acts on behalf of, or instead of, a human. The access-control obligation under Article 21 is not satisfied by demonstrating that human users are correctly authorized. It is satisfied by demonstrating that all actors, human and non-human, operate within bounded, reconstructable authority.

Article 23 (Incident notification) imposes a 24-hour early warning, a 72-hour notification, and a final report within one month. The timeline assumes the entity can characterize the incident within those windows. For agent-mediated incidents, characterization requires answering the reconstruction question: which authorization chain produced the affected action, what scope held at each hop, and how far the compromise propagated through the delegation tree. Without architectural reconstruction, the 24-hour window may pass before the entity can even identify the chain involved.

Annex I and Annex II (Sectoral entity classifications) define the scope of obligation. Critical infrastructure entities in energy, water, transport, and digital infrastructure carry heightened obligations under Article 21. The agent fleets deployed in those sectors (OT and SCADA supervisory agents, grid-management agents, transport orchestration agents) sit precisely in the Article 21 access-control intersection where retrofit instrumentation fails most visibly.

Evidence artifact required: a reconstructable access-control trail for non-human actors, with cross-organizational chain integrity for supply-chain incidents under Article 21 paragraph 2(d) on supply chain security.

How the architecture satisfies: the receipt chain at every delegation hop is the access-control trail by construction. The out-of-chain revocation log (Paper #4 pattern) is the post-incident reconstruction artifact that survives partial-chain compromise. Zero-knowledge verification (Paper #3) applies where access-control evidence

must be produced without disclosing the underlying data, common in supply-chain incidents where confidentiality obligations bind multiple parties.

Sector application: critical infrastructure agent fleets. The recurring scenario is a supervisory-layer agent that delegates to a device-level agent, which delegates to a firmware-update agent. Without strict attenuation discipline across the chain, a partial compromise at any hop produces an incident the entity cannot characterize within Article 23's 24-hour window. With the architecture, the receipt chain produces the characterization automatically, and the cross-enclave attenuation pattern from Paper #4 keeps the chain coherent across security boundaries.

DORA: financial services and ICT third-party risk

DORA (Regulation EU 2022/2554) applies to financial entities (banks, investment firms, insurers, crypto-asset service providers, market infrastructure operators) and their critical ICT third-party service providers. The regulation has applied since January 17, 2025. Joint technical standards from the European Banking Authority, the European Securities and Markets Authority, and the European Insurance and Occupational Pensions Authority give the regulation its operational shape.

The articles that map directly to agent deployment:

Article 5 (ICT risk management framework requirements) establishes the principle that the ICT risk management framework must be documented and reviewed at least annually. For institutions deploying AI agents, the documentation must include the agent risk surface, which means the framework must be able to describe agent authority bounds in a form the regulator can interpret.

Article 6 (ICT risk management framework, the documented framework itself) requires a clear allocation of roles and responsibilities for all ICT-related risks. Agent-mediated risks fall under this allocation. Where agent delegation chains cross organizational boundaries, the framework must document how authority bounds are preserved across those boundaries.

Article 8 (Identification and mapping) requires identification and classification of ICT-supported business functions, information assets, ICT assets, roles, dependencies, and ICT risk. Agent systems are ICT assets. Agent delegation chains are dependencies. The mapping obligation is not satisfied by an asset inventory that lists agents as opaque black boxes. It is satisfied by a mapping that includes the authorization graph and the chain structure.

Article 9 (Protection and prevention measures) requires technical and organizational measures to ensure the resilience, continuity, and availability of ICT systems and to maintain high standards of security, confidentiality, and integrity. For agent systems, the integrity obligation extends to the integrity of the authorization chain.

Articles 17 through 19 (ICT-related incident management) require classification, reporting, and post-incident review of major ICT-related incidents. Article 18 sets the classification criteria for major incidents (clients and transactions affected, duration, geographical spread, data losses, criticality of services, economic impact). Article 19 then drives the reporting flow: initial notification within a few hours of classification (and within twenty-four hours of detection), intermediate report at seventy-two hours, and final report at one month, with root-cause analysis required in the final report. For agent-mediated incidents, root-cause analysis is the reconstruction question, and reconstruction in turn is what the receipt chain produces by construction.

Articles 24 through 27 (Digital operational resilience testing, including TLPT) require regular testing of ICT systems, with the most advanced testing under the threat-led penetration testing (TLPT) framework applying to selected significant financial entities. Agent systems are now in scope as both attack surface and attacker

capability; TLPT under the advanced testing framework increasingly simulates delegation-chain escalation as a real-world attack pattern. (DORA Article 23, by contrast, addresses operational and security payment-related incident reporting and is not the TLPT article, despite the colloquial conflation common in early commentary.)

Articles 28 through 30 (ICT third-party risk management) require contractual arrangements, monitoring, and exit strategies for ICT third-party services, including subcontracting chains. Agent delegation chains that traverse third-party services bring the third-party risk obligation directly into the architecture.

Evidence artifact required: operational-resilience evidence that ICT-supported functions, including non-human actor activity, can be reconstructed end-to-end. Explicit identification of third-party ICT dependencies in the chain. Demonstrable capacity to terminate exposure under Article 28 when contractual or risk-based grounds require it.

How the architecture satisfies: the receipt chain is the operational-resilience evidence artifact. Cross-organizational delegation receipts identify third-party participation in the chain by construction; no separate inventory exercise is needed at incident time. Revocation that lives outside chain integrity (Paper #4 invariant 3) satisfies the Article 28 obligation to terminate exposure even when a third-party intermediate is uncooperative or compromised.

Sector application: financial-services trading agents, algorithmic execution agents, risk-adjusted decision agents, customer-service AI agents. The recurring scenario is a portfolio-manager authority delegated to an algorithmic-execution agent, then to a smart-order-router, then to a venue-specific micro-strategy. The Edinburgh framework (Szpruch et al., April 2026) describes the policy layer of this delegation graph; Papers #2 through #4 describe the protocol layer that makes the policy enforceable in chains. The architecture covers both layers for the financial-services regulator.

EU AI Act: high-risk AI systems

The EU AI Act (Regulation EU 2024/1689) classifies AI systems by risk tier. High-risk systems under Annex III include AI used in employment, education, law enforcement, migration, justice, democratic processes, critical infrastructure operation, essential private and public services (including credit scoring), and safety components of products. The regulation is scheduled to become applicable on 2 August 2026 with exceptions. Per the 7 May 2026 AI Omnibus political agreement, stand-alone high-risk AI systems are expected to apply from 2 December 2027, and high-risk AI systems embedded in regulated products (Annex II coverage) from 2 August 2028. The window for architectural alignment is therefore staged: DORA and NIS2 are binding today; the AI Act high-risk regime is expected to bind from late 2027 onward.

The articles that map directly to agent deployment:

Article 9 (Risk management system for high-risk AI systems) requires a documented, iterative risk management process across the AI system lifecycle. For agent systems, the risk surface includes the authorization architecture; the risk management documentation must describe how chain integrity is preserved.

Article 10 (Data and data governance) requires that training, validation, and testing data sets meet specified quality criteria, including representativeness and bias mitigation. Where agent systems access personal data through delegation chains, the data-governance obligation extends to whether the chain narrows scope to the minimum necessary at each hop. Attenuation discipline aligns directly with the data-minimization principle.

Article 13 (Transparency and provision of information to deployers) requires that providers ensure the AI system is sufficiently transparent for deployers to interpret and use the output appropriately. For agent systems, transparency includes the agent's authorization model. Deployers who cannot inspect the chain cannot satisfy their own deployer-side obligations under Article 26.

Article 14 (Human oversight) requires that high-risk AI systems be designed and developed in such a way that they can be effectively overseen by natural persons during the period in which the AI system is in use. The often-quoted phrase about "natural persons" is sometimes read as requiring real-time human-in-the-loop. The architectural reading is different and more defensible: human oversight is satisfied asymptotically when natural persons can interpret and override the agent's decisions after the fact, based on a receipt chain they can read. The chain is the oversight artifact. Without the chain, oversight collapses into either real-time human approval of every action (operationally infeasible at agent scale) or operator self-attestation (regulator-unacceptable under Annex IV evidence requirements).

Article 16 (Provider obligations) sets the core duties of providers of high-risk AI systems: conformity assessment, technical documentation under Article 11, risk management, record-keeping, and registration. Providers are the entities that develop or substantially modify a high-risk AI system and place it on the EU market under their own name.

Article 17 (Quality management system) requires the provider to establish, document, and maintain a quality management system covering compliance with applicable AI Act requirements, including design control, validation, change management, and post-market monitoring procedures.

Article 72 (Post-market monitoring) obligates providers to set up and document a post-market monitoring system proportionate to the AI system's risks. The system must collect, document, and analyze data on AI system performance throughout its lifetime.

Article 73 (Serious incident reporting) requires providers to report serious incidents to market surveillance authorities within prescribed timelines, with shorter windows for life-threatening malfunctions and widespread infringement.

Provider versus deployer distinction. Annex IV technical documentation is primarily a provider obligation under Article 11. Deployers under Article 26, including most financial-services and critical-infrastructure organizations using third-party AI systems, must operationalize the documentation, logs, and oversight evidence the provider supplies, rather than produce the documentation themselves. Where a regulated organization develops or substantially modifies an AI system internally (a pattern increasingly common in financial services and critical infrastructure where proprietary advantage matters), the organization assumes the provider role and inherits the full Article 16 / 17 / 72 / 73 obligation stack alongside Annex IV. The architectural answer in this paper applies to both postures: providers produce the receipt-chain evidence by construction; deployers operationalize it as part of their Article 26 oversight duties.

Annex IV (Technical documentation) is the most operationally significant. It enumerates required documentation items including: - Item 2(c): a description of the architecture, including how the system components build on each other or feed into each other - Item 2(g): a description of validation and testing procedures, including pre-determined performance metrics - Item 2(h): a description of cybersecurity measures - Item 3: detailed information about the system's monitoring, functioning, and control - Item 4: a description of the performance metrics - Item 5: a description of the risk management system

Annex IV item 2(c) maps directly onto the AAC architecture diagrams and the delegation-discipline description. Item 2(h) maps onto the cybersecurity primitives: capability tokens, zero-knowledge verification, and attenuation discipline are themselves the cybersecurity measures. Item 3 maps onto the receipt-chain queryability: monitoring is not a separate dashboard, it is a property of the architecture.

Evidence artifact required: decision-provenance reconstruction for high-risk AI system actions, sufficient for a natural-person auditor to interpret and override.

How the architecture satisfies: Annex IV item 2(c) is satisfied by AAC architecture documentation and the delegation-discipline diagrams. Item 2(h) is satisfied by the cybersecurity primitives composed across Papers #2, #3, and #4. Article 14 human oversight is satisfied asymptotically through the receipt chain as the oversight artifact, resolving the apparent tension between agent-scale autonomy and natural-person interpretability.

Sector application: high-risk AI systems across Annex III categories. Recurring scenarios include healthcare diagnostic agent chains (patient-data agent to diagnostic-service agent to research aggregator), HR and employment agent chains (CV-screening agent to decision-recommendation agent to communication agent), and critical-infrastructure operation agents (already mapped under NIS2 above, applicable in parallel under EU AI Act when classified as Annex III).

Cross-regime convergence

Three regulations name different obligations under different enforcement regimes. They converge on the same artifact.

NIS2 Article 21 access-control reconstruction, DORA Articles 17 through 19 incident reconstruction, and EU AI Act Annex IV decision-provenance are not three audit programs. They are three queries against the same architectural property: a receipt chain produced by attenuation discipline that survives partial-chain compromise. The receipt chain is regulator-readable by design, interpretable by auditors from each regime without translation.

The cost of fragmentation is real. Enterprises currently running three parallel audit programs duplicate effort by a factor of three to five, depending on the depth of cross-functional coordination already in place. Architecture-by-construction collapses the cost, because the evidence artifact is the same artifact under all three queries. The compliance team does not produce three different reports; the regulator does not request three different reconstructions. Both sides query the same chain.

This convergence is not coincidence. The three regulations were drafted independently to address different policy concerns. They converged on the same evidence requirement because the underlying problem (reconstructing non-human-actor authority across delegation chains) admits the same architectural solution regardless of which regulatory framing initiates the query. Parallel discovery across regulators is a signal of architectural stability, in the same way that parallel discovery across academic disciplines signals the convergence of a mathematical structure.

Part III: Patterns and sector applications

Three patterns that satisfy audit

1. **Receipt-chain as primary evidence artifact.** Every agent action has cryptographic provenance bound to its originating authority. The auditor queries the chain by action ID; the chain produces the answer without operator cooperation. AAC architecture A2A flow steps 13 through 15 implement this pattern; Paper #4 describes the discipline that makes the chain reliable in production.

2. **Out-of-chain revocation log as ICT-incident-response artifact.** Transparency log or epoch-anchored revocation list. Survives partial-chain compromise. Satisfies DORA Article 17-19 incident-management obligations and NIS2 Article 23 notification timelines simultaneously. The architecture is the same; the regulator-side reading is different per regime.

3. **Asymmetric audit: regulator queries the chain, operator does not curate.** The architecture is operator-independent. The regulator validates the chain without the operator's narrative. This is the structural difference between architecture-as-evidence and instrumentation-as-evidence, and it is the property regulators have begun explicitly requesting under all three regimes.

Three patterns that fail

1. **Application logs as audit.** Operator-attested, retrofit, not authorization-bound. Cannot answer the reconstruction questions in Part I. Most enterprise audit programs as of 2026 still rely primarily on application logs; the gap will become explicit under regulator examination within the next twelve to eighteen months.

2. **Single-system reconstruction.** Breaks at chain composition. Cross-organizational, cross-cloud, and cross-domain delegation chains require cross-organizational receipt assembly, which single-system instrumentation cannot produce. The composition failure is invisible to internal audit until an external examination surfaces it.

3. **Self-attestation.** Operator says "we comply" with no architectural evidence to anchor the claim. Under NIS2, DORA, and the AI Act, attestation is increasingly expected to be backed by evidence artifacts that survive operator-independent inspection where the regulated question is reconstructability of authority, control, or incident root cause. Compliance posture that rests solely on policy attestation without architectural evidence is increasingly fragile under regulator examination.

Sector deep-dives

Financial services under DORA. The recurring scenario is the algorithmic-trading delegation chain. A portfolio-manager authority is delegated to an algorithmic-execution agent, which delegates to a smart-order-router, which delegates to a venue-specific micro-strategy. Each delegation is correct in isolation; the composed chain can re-anchor the portfolio's risk budget without re-authorization unless attenuation discipline holds at every hop. The Edinburgh framework (Szpruch et al., April 2026) describes the policy layer for this scenario; Papers #2 through #4 describe the protocol layer that makes the policy enforceable. The full architecture covers both layers for the DORA examination.

Critical infrastructure under NIS2. The recurring scenario is the multi-domain command chain in OT and SCADA environments. A supervisory-layer agent delegates to a device-level agent, which delegates to a

firmware-update agent. If the firmware-update agent runs in a different security enclave than the supervisory layer, the cross-domain chain break manifests: caveats bounding the supervisory authorization may not transfer to the firmware enclave's policy framework. The Paper #4 cross-enclave attenuation pattern is the architectural answer. NIS2 Article 21 access-control reconstruction becomes a chain-walk against the receipt artifact at incident time, instead of a coordination exercise across operational technology and information technology teams under time pressure.

High-risk AI under the EU AI Act, healthcare example. The recurring scenario is cross-institution patient-data delegation. A patient consents to data sharing with a specific provider. The provider's agent delegates to a third-party AI diagnostic service. The diagnostic service delegates to a research aggregator. Each delegation may be compliant in isolation under GDPR Article 9 special-category data rules. The composed chain, without strict attenuation, can violate the patient's original consent scope by the time the data reaches the research aggregator. EU AI Act Annex IV decision-provenance reconstruction requires that the architecture can show, by chain integrity, where consent scope held and where it did not. The cross-domain attenuation pattern from Paper #4 is the architecture that makes the demonstration possible.

The three sector applications share a common shape. The regulatory obligation is reconstruction. The architectural answer is the receipt chain. The pattern that fails in each sector is the same pattern: operator-attested logs without chain integrity, composed across delegation hops where the composition was never audited as such.

What to do Monday morning

For CISOs, Chief AI Governance Officers, DPOs, and audit committee leads in regulated organizations deploying AI agent systems:

1. **Inventory.** Which AI agent systems are deployed in production, and which regulations apply? Run the sectoral check under NIS2 Annex I and II, the scope check under DORA, and the high-risk classification check under EU AI Act Annex III. Most organizations are in scope of at least one; many discover they are in scope of two or three when they look carefully.

2. **Gap analysis.** What does each regulation demand against what the current architecture produces? Specifically, for each agent-mediated action class, can your team reconstruct the originating authorization, the scope narrowing at each delegation hop, and the revocation propagation path? If any of the three answers is no, the gap is architectural and cannot be closed by additional logging alone.

3. **Architecture move.** Capability tokens with hash-bound caveats. Receipt chain at every delegation hop. Revocation outside chain integrity. Three invariants from Paper #4, three regulations covered in this paper, one architectural answer. The implementation effort scales with the number of agent platforms in scope, not with the number of regulations; the architecture solves the cross-regime convergence problem in addition to each regime in isolation.

4. **Audit-readiness check.** Can you reconstruct an arbitrary agent action's authorization chain end-to-end, today, in less than twenty-four hours? Can you produce the chain to a regulator on notice without operator narrative curation? Run the simulation against a real action, not a designed test case. If the simulation fails, the gap is architectural and the remediation begins with the receipt-chain layer.

5. **Cross-disciplinary brief.** Bring legal, DPO, CISO, audit committee, and architecture into one room. The regulatory landscape requires all five lenses; siloed compliance programs cannot answer the cross-regime

convergence question. The brief is not a one-time exercise. It is the basis for the ongoing compliance posture under regulations whose enforcement intensity is climbing through 2026 and 2027.

Closing Series I

Paper #5 closes the Non-Human Identity series.

The five papers in order:

1. **Agents Are Not Service Accounts.** The identity primitive: agents need first-class non-human identity, not retrofit service-account credentials.
2. **Authorization for AI Agents: Beyond RBAC.** The authorization model: capability tokens at single-hop evaluation, intersection authority across user, agent, and task.
3. **Authorization Without Disclosure: Zero-Knowledge Proofs for Agent-to-Agent Authorization.** Verification without disclosure, the privacy-preserving layer that makes capability verification defensible across trust boundaries.
4. **Delegation Without Escalation.** Attenuation discipline at every hop: strict-subset scope narrowing, time-bounded forward chains, revocation outside chain integrity.
5. **Auditing Agents Under NIS2, DORA, and the EU AI Act.** Regulator-readable evidence by construction, the same architecture satisfying three regulatory regimes through a single receipt-chain artifact.

The architecture is complete in five steps. Identity, authorization, verification, delegation, audit. Each layer composes with the next; each layer is necessary, none is sufficient alone. The series describes a non-human identity architecture that holds under deep delegation chains, across organizational boundaries, under adversarial regulatory examination.

The next layer is the substrate. Series II opens in late June 2026, addressing post-quantum cryptography for AI agent systems: the cryptographic primitives that the authorization, verification, delegation, and audit architecture rests on, and what must change in those primitives as the post-quantum migration window opens. The compliance window that closes for non-human identity in 2026 and 2027 is the same window that opens for post-quantum readiness across the same regulatory regimes. The architecture for both is continuous.

Acknowledgments

Thanks to Agustin Martinez Fayo for the Paper #4 v1.1 peer review on revocation framing that carried forward into the audit-readiness chain reading here. Thanks to Cesar Cerrudo for the conversations on substrate-layer review that shape the bridge to Series II. Thanks to the ECA and to Dominique Tessier for the institutional grounding of the substrate-sovereignty thesis that frames Series II’s opening question.

Citations and related work

Series I prior papers (each carries a permanent DOI on Zenodo; cite by version DOI for this exact text or by concept DOI for "latest version"):

- **Paper #1 of this series.** *Agents Are Not Service Accounts: Why Non-Human Identity Needs Its Own Model.* DOI 10.5281/zenodo.20242385. The identity primitive: agents need first-class non-human identity, not retrofit service-account credentials.
- **Paper #2 of this series.** *Authorization for AI Agents: Beyond RBAC. Capability Tokens, User-Context-Aware Policy, and the Limits of Roles.* DOI 10.5281/zenodo.20242387. Capability tokens at single-hop authorization layer, intersection authority across user, agent, and task.
- **Paper #3 of this series.** *Authorization Without Disclosure: An advisory reference architecture for non-disclosing agent-to-agent authorization.* DOI 10.5281/zenodo.20242389. Zero-knowledge verification of capability decisions across trust boundaries.
- **AAC Construction Specification.** Technical companion to Paper #3, deep protocol detail referenced across the series. DOI 10.5281/zenodo.20242391.
- **Paper #4 of this series.** *Delegation Without Escalation: Capability tokens, attenuation discipline, and the patterns that survive in production.* DOI 10.5281/zenodo.20242393. Attenuation discipline at every hop: strict-subset scope narrowing, time-bounded forward chains, revocation outside chain integrity. The receipt-chain artifact this paper relies on.

External works carried forward from Papers #1 through #4:

- **Birgisson et al., 2014.** *Macaroons: Cookies with Contextual Caveats for Decentralized Authorization in the Cloud.* ACM CCS. The foundational primitive underlying attenuation discipline.
- **Edinburgh Szpruch et al., April 13, 2026.** *Scalable Runtime Governance for Agentic AI in Financial Services.* Policy-layer framing for capability decomposition under DORA; complementary to this paper's protocol-layer treatment.
- **WIT-SVID (spiffe/spiffe#362).** In-flight per-agent SPIFFE identity primitive, application-layer holder-of-key binding. https://github.com/spiffe/spiffe/pull/362
- **draft-ietf-wimse-arch.** IETF WIMSE WG architecture draft. Section 3.3.9 on multi-hop delegation chains. https://datatracker.ietf.org/doc/html/draft-ietf-wimse-arch
- **draft-ietf-oauth-spiffe-client-auth.** IETF OAuth WG. SVIDs as OAuth client credentials. https://datatracker.ietf.org/doc/html/draft-ietf-oauth-spiffe-client-auth
- **AAC Construction Specification.** Companion document, deep protocol detail for the architecture referenced across the series.
- **OWASP Top 10 for Agentic Applications 2026.** ASI03 Identity and Privilege Abuse (primary mapping), ASI09 Insecure Logging and Monitoring, ASI10 Rogue Agents.
- **CSA AI Security Maturity Model (AISMM) v3.7, May 7, 2026.** Five-level maturity framework, twelve domains. Privacy and Compliance domain assessment from Level 3 (Defined) and above is satisfied by the receipt-chain evidence model.
- **Grantex State of AI Agent Security 2026.** Audit of 30 popular open-source AI agent projects. Primary source for the agent-framework identity gap.
- **Gravitee State of AI Agent Security 2026.** 45.6% shared agent-to-agent credentials in surveyed deployments.

- **OECD.AI incident 2026-05-04-4a73.** Grok / Bankrbot prompt-injection exploit, May 4, 2026.

New for Paper #5:

- **NIS2 Directive (EU 2022/2555).** Full text via EUR-Lex. ENISA implementation guidance and national transposition references for member-state enforcement specifics.
- **DORA Regulation (EU 2022/2554).** Full text via EUR-Lex. Joint technical standards from EBA, ESMA, and EIOPA for operational definitions of Article 6, 8, 17-19, and 28 obligations.
- **EU AI Act (Regulation EU 2024/1689).** Full text via EUR-Lex. EU AI Office implementing acts on Annex IV technical documentation and Article 17 provider obligations.
- **ENISA Report on AI Cybersecurity Challenges, 2023 with 2024 and 2025 updates.** Sectoral threat-landscape reference for NIS2 Article 21 risk-management measure design.
- **GDPR Regulation (EU 2016/679) Article 9.** Special-category data processing, cross-referenced for healthcare and HR sector deep-dives.

Paper #5 of 5 in the Non-Human Identity series. Series I closes here. Series II opens late June 2026 with the post-quantum substrate question.